Explainable Artificial Intelligence

Michael Ridley, Librarian Emeritus, University of Guelph; PhD Candidate, Western University; and Postgraduate Affiliate, Vector Institute

Introduction

Algorithmic decision-making, enabled by machine learning, is ubiquitous, powerful, often opaque, sometimes invisible, and, most importantly, consequential. Machine learning is embedded in many information tools and systems, central to numerous research methods, and pervasive in the applications of everyday life. Safiya Noble emphasizes the critical nature of artificial intelligence (AI) by observing that it will become "a major human rights issue in the twenty-first century."[1] As with nearly all aspects of contemporary life, AI is having a profound influence on research libraries, scholarly communication, and key functions of the academy.

Because "authority is increasingly expressed algorithmically,"[2] it is crucial that this authority be interrogated and assessed with the same rigor and appropriate methods relevant to all aspects of the academic mission. Machine learning and deep learning are potent technologies that will be utilized to great advantage. However, "the danger is not so much in delegating cognitive tasks, but in distancing ourselves from–or in not knowing about–the nature and precise mechanisms of that delegation."[3] Hence the critical importance of "explainable artificial intelligence" (XAI) and its two pillars: trust and accountability.

XAI is a diverse set of strategies, techniques, and processes that render AI systems interpretable and accountable. While some XAI approaches are highly technical, involving the perturbation of individual features in multi-layer neural network models, others are broad social and political policies enacted through regulation or legislation. Whatever the approach, XAI emphasizes explainability as an essential requirement for a technology that has for too long been defined by its opacity and what Frank Pasquale calls "remediable incomprehensibility."[4]

The use and development of machine learning applications in research libraries will only continue to grow in volume and influence. As AI reconfigures much of scholarly communications, it will be essential that libraries, and their users, have trust in the cognitive delegation of many tasks and processes. Mariarosaria Taddeo notes that "delegation without supervision characterises the presence of trust."[5] Approaching that state will require artificial intelligence to exhibit, and be open to, new levels of transparency and accountability. One critical element of that is explainability.

"As AI reconfigures much of scholarly communications, it will be essential that libraries, and their users, have trust in the cognitive delegation of many tasks and processes."

Defining Explainable AI (XAI)

The US Defense Advanced Research Projects Agency (DARPA) definition of XAI is widely referenced. The purpose of XAI is to enable human users "to understand, appropriately trust, and effectively manage the emerging generation of artificially intelligent partners."[6] To this user-centric XAI definition, DARPA adds the expectation that AI systems "will have the ability to explain their rationale, characterize their strengths and weaknesses, and convey an understanding of how they will behave in the future."[7] Examining these definitions yields both insights and complications.

The user-centric definition has three key concepts: understanding, trust, and management. Understanding can mean a range of ideas from simple awareness or acceptance to acknowledgement and finally to detailed knowledge. While the idea of trust seems straightforward, the modifier "appropriately" suggests a conditional situation where the granting of trust is contextual. Managing indicates a relationship between the user and the AI and implies that the human user is, or should be, in a position of reasonable control. However, referencing AI systems as "partners" suggests a more cooperative and quasi-independent relationship.

The system-centric definition also has three key concepts: rationale, strengths and weaknesses, and future behavior. The rationale could pertain to the purpose of the AI, the logic of its model, a justification for its actions, or its application in specific situations. The disclosure of strengths and weaknesses indicates a level of openness and transparency that would make obvious system limitations and key assumptions. It also seems likely to conflict with trade secrecy, intellectual property issues, and data privacy. The emphasis on future behavior recognizes that AI will be an ongoing part of everyday life, hence the need for predictability and consistency. It also implies that AI will be subject to longitudinal evaluations to ensure levels of performance.

European Union General Data Protection Regulation

It is difficult to overestimate the impact of the European Union (EU)'s 2018 General Data Protection Regulation (GDPR) on XAI. The GDPR's "right to explanation" regarding algorithmic decisions is having a global reach (the so-called "Brussels effect"), causing debate and regulatory review well beyond the EU. While interpretability has always been a concern in computer science, the GDPR has refocused this issue as an explainability problem and made it a public policy question.

The explanatory requirements in the GDPR are actually quite narrow, but their impact has been much broader, with jurisdictions as diverse as Canada and the City of New York developing impact assessment protocols with respect to algorithmic decisions that include explainability requirements. As seen with the "right to be forgotten," international legislation or regulation can have a profound effect on national affairs. The global nature of digital technologies is a reminder that monitoring the policy agendas of other jurisdictions is important.

Opacity and Trust

Why do we need an explanation for how AI works? Geoffrey Hinton, often referred to as the godfather of deep learning and neural networks, observes, "A deep-learning system doesn't have any explanatory power....the more powerful the deep-learning system becomes, the more opaque it can become."[8] Despite this, Hinton has been critical of requirements that AI should explain itself and insists performance should be the key measure of trust. After all, humans can't provide explanations for many of their actions or decisions, why expect AI to do otherwise?

While Hinton may discount the importance of, or even the need for, an explanation, psychologists and cognitive scientists do not. Explanations are "more than a human preoccupation—they are central to our sense of understanding, and the currency in which we exchange beliefs."[9] There is an extensive literature on both the power and the failings of AI. Examples of discrimination and unfairness are matched by extraordinary advances and success. However, it is exactly for these reasons that the opacity, complexity, and consequential nature of AI drives the need for trust and elevates explanation as a key antidote.

> "As academic libraries increasingly acquire and develop algorithmic decision-making systems and services in support of scholarly communications and the operation of the library, they must do so in a manner that insists on interpretability and explanation."

What is a good or satisfactory explanation? For whom is the explanation provided, in what context, with what, if any, evidence, and presented in what manner? An explanation should be able to address "how" (inputs, output, process), "why" (justification, motivation), "what" (awareness that an algorithmic decision-making system exists), and the "objective" (design, maintenance)."[10] In the context of opaque systems, an explanation should be:

> "(1) model-agnostic, so it can be applied to any black box model;
>
> (2) logic-based, so that explanations can be made comprehensible to humans with diverse expertise, and support their reasoning;
>
> (3) both local and global, so it can explain both individual cases and the overall logic of the black-box model;
>
> (4) high-fidelity, so it provides a reliable and accurate approximation of the black box behavior."[11]

A more holistic view would include explanations that consider the data used for training and decision-making, the computational environment utilized, the context of the algorithmic design and deployment, and those responsible for its operation and use (that is, a sociotechnical analysis).[12] Technical explanations are required for those involved in system design and performance testing while accessible explanations are needed for those affected by algorithms. In the latter context, a good explanation is contrastive ("why P not Q?"), selective (only certain evidence is required not a complete explanation), and social (a dialogue, interactive, contextual).[13]

As academic libraries increasingly acquire and develop algorithmic decision-making systems and services in support of scholarly communications and the operation of the library, they must do so in a manner that insists on interpretability and explanation. To do anything less is an unknowing delegation to technology and an abrogation of scholarly rigor.

XAI Strategies, Techniques, and Processes

Approaches to XAI can be broadly categorized as proofs, validations, and authorizations. Within these categories are numerous explanatory practices, which are contextual, system or model dependent, and audience specific.

Proofs

Proofs as explanations are testable, demonstrable, traceable, and unambiguous. In the context of AI, they pertain primarily to rule-based expert systems or systems that use decision trees (that is, an explicit knowledge basis encoded in human interpretable statements). Proofs of algorithmic predictions require clear causal links and logic statements that unambiguously trace performance from data to decision. Such an examination is possible in AI systems that employ ruled-based or decision-tree models because the rationale is specifically coded and human readable. While the performance of rule-based and decision-tree systems is inferior to that of current machine-learning techniques, these systems are still in use where explicit causality and accountability can be documented (for example, in certain health care, insurance, and public sector applications), demonstrating that, in specific circumstances, explainability is preferred over performance.

Validations

Validations or verifications as explanations are conclusions about the veracity of the AI substantiated by evidence and/or reason. Verification confirms the AI performance against an external measure, standard, factual data, or third-party corroboration.

Feature selection is an explanatory strategy that attempts to reveal the key factors (for example, hyperparameter weights) that had a primary role in the prediction of the algorithm. By isolating or adjusting these elements, it is possible to explain the key components of the decision. There are various feature selection techniques but all of them are "decompositional" in that they attempt to reduce the work of the algorithm to its component parts and then use those results as an explanation. Feature selection provides a verification that certain elements have a primary influence on the prediction thereby explaining why a certain outcome pertains but not another (a contrastive explanation). While such an explanation is used primarily for designers to adjust their models (that is, it is an error-correction process), allowing users to examine feature selection explanations would provide a justification for why a decision was made and would allow them a basis to challenge that result.

In seeking explanations, people rarely ask for or rely on complete explanations. Rather than reviewing and assessing all the causes (even if provided), people tend to be highly selective. We seek and accept explanations that "satisfice." Approximation or abstraction are techniques that create a more simplified model to explain the more complex model. Approaches such as model distillation or model-agnostic feature reduction create a simplified presentation of the algorithmic model. This approximation or abstraction may compromise accuracy, but it provides an accessible representation that enhances understandability.

XAI researcher Trevor Darrell believes that "the solution to explainable AI is more AI."[14] In this approach to explanation, oversight AI are positioned as intermediaries between an AI and its users. These AI have been called "ethical governors,"[15] "flight data recorders,"[16] and, more ominously, "AI guardians."[17] These examples of intelligent middleware offer the ability to interpolate the values and expectations of third parties, such as research libraries, in the process of deriving an explanation from an AI.

Replication is a recognized verification strategy in many aspects of research. Being able to independently reproduce results in different settings provides evidence of veracity and supports user trust. However, documented problems in reproducing machine-learning research have questioned the generalizability of these approaches and undermined their explanatory capacity. In response, a "Reproducibility Challenge" was created by the International Conference on Learning Representations (ICLR) to validate 2018 and 2019 conference submissions.[18] More rigorous replication through the availability of all necessary components will be important to this type of verification.

Authorizations

Authorizations as explanations are processes, typically involving third parties, which provide an assessment or ratification of the AI. Authorizations might pertain to the AI model, its operation in specific instances, or even the process by which the AI system was created. Examples of authorization include transparency, expertise, due process, litigation, and liability. This section will look at voluntary codes, audit, legislation, and regulation.

Voluntary codes or standards that encourage explanatory capabilities are approaches to explanation supported by the AI industry and professional organizations (for example, Association for Computing Machinery and IEEE). Self-regulation through non-binding codes or standards is a type of governance that some argue is the most effective for rapidly changing technologies. The inflexibility of legislation and regulation might either unnecessarily constrain AI or be ineffective in managing new developments. The "privacy by design" initiative might be a model for something like "explanation by design" whereby prior impact assessment reports, certification requirements, and codes of conduct would provide incentives for more "scrutable algorithms." Unfortunately, this strategy is undercut by the poor experience with voluntary mechanisms regarding privacy protection.

A commonly recommended approach to AI explanation is third-party auditing. The use of audits or audit principles is widely accepted in a variety of areas. While auditing is typically *ex post*, it can be accomplished at any stage, including design specifications, completed code, operating models, or periodic audits of specific decisions. Auditing for XAI would require trusted auditors, an accepted set of standards to measure against, and the "auditability" of the algorithms or systems. Critics of the audit approach have focused on lack of auditor expertise, algorithmic complexity, and the need for *ex ante* approaches.

The efficacy, and likelihood, of legislation mandating explanatory AI is widely discussed among researchers. While US, and to a lesser extent Canadian, past practice signals a reluctance to legislate in these areas, the EU, France, and the United Kingdom are taking different and more proactive approaches as exemplified by the GDPR. As a result, in Canada and the US, the most common recommendation for AI oversight and authorization is the use of a regulatory agency. Such an agency would have legislated or delegated powers to investigate, certify, license, and arbitrate on matters relating to AI and algorithms, including their design, use, and effects. The breadth and depth of the responsibilities of these agencies varies by those promoting them and by the relevant jurisdiction. Specific suggestions for a public agency include a "neutral data arbiter" with investigative powers like the US Federal Trade Commission, a Food and Drug Administration "for algorithms," a standing "Commission on Artificial Intelligence," quasi-governmental agencies such as the Council of Europe, and a hybrid model combining certification and liability. There are few calls for an international regulatory agency despite the global reach for many, if not most, AI systems and services.

XAI and Research Libraries

Algorithmic decision-making is already pervasive in information tools and services acquired, provided, or developed by research libraries. Often the methods and processes of those tools and services are invisible or unacknowledged. If libraries are to trust the quality, value, and credibility of these innovations, it is important that they be explainable.

David Lankes warns of a new digital divide with "a class of people who can use algorithms and a class used by algorithms,"[19] and argues that "librarians need to become well versed in these technologies, and participate in their development, not simply dismiss them or hamper them. We must not only demonstrate flaws where they exist but be ready to offer up solutions. Solutions grounded in our values and in the communities we serve."[20] This is echoed by Catherine Coleman in her assertion that librarians can be co-creators of "an intelligent information system that respects the sources, engages critical inquiry, fosters imagination, and supports human learning and knowledge creation."[21] There are numerous examples, such as Hamlet from MIT, the AI Lab at the University of Rhode Island, and the Stanford Library AI initiative, where machine learning in research libraries is occurring with an emphasis on explainability and accountability.[22]

Developments such as these highlight Chris Bourg's 2017 suggestion that "we would be wise to start thinking now about machines and algorithms as a new kind of patron."[23] In doing so, research libraries need to consider not merely how the data can be exposed to algorithmic systems, but the new obligations with respect to data privacy and reuse that may come from this. These implications may extend beyond what is currently considered in research data management protocols.

> Algorithmic decision-making is already pervasive in information tools and services acquired, provided, or developed by research libraries.

An illustration of why research libraries need to accelerate their involvement in AI and XAI arises from a recent breakthrough in the unsupervised text mining of the scientific literature, which demonstrated "that latent knowledge regarding future discoveries is to a large extent embedded in past publications."[24] This insight was observed previously during the formative years of Medline[25] and has motivated the current "knowledge validation engine" of Project Aiur from Iris.ai.[26] Each of these projects acknowledges that the structure of scientific communications (for example, the nature of abstracts) enables machine-learning analysis and highlights the need to verify the outcomes by examining the processes. They also emphasize the challenges of explainability when the research literature is being utilized and interpreted using complex and often opaque methods.

It is concerning that these innovations are occurring outside the field of academic librarianship and with little or no involvement of library expertise. If libraries are to shape AI development and embed values such as explainability in these tools and services, it is essential that the challenges voiced by Lankes, Bourg, and Coleman be acknowledged, accepted, and acted upon. In addition to the focus on innovation in tools and services, academic libraries can further XAI through such avenues as public policy and algorithmic literacy.

> If libraries are to shape AI development and embed values such as explainability in these tools and services, it is essential that the challenges voiced by Lankes, Bourg, and Coleman be acknowledged, accepted, and acted upon.

Public Policy

A key XAI strategy is to use authorizations, such as legislation, regulation, and audit, as governance methods to support, or even require, explainability. Despite widespread concerns about algorithmic decision-making with respect to bias, discrimination, and unfairness, this is an area that is largely unregulated in Canada and the United States. The AI public policy landscape is nascent. Some have argued for a "regulatory lag" to allow more clarity on how AI will evolve, while others more cynically dismiss all regulations as solving "yesterday's challenges" and impeding innovation in a globally competitive "AI race." While premature and reactive regulation is undesirable, neither is an environment where abuses, harms, and predatory practices are allowed to exist.

Research libraries, through organizations such as the Association of Research Libraries and the Canadian Association of Research Libraries, have a strong interest in influencing public policy and have achieved substantial successes in this area, even if only in raising public awareness. While it is argued that blanket AI regulation will be less effective than application-specific regulation (for example, let those who regulate air travel regulate AI in air travel), there are overarching principles, such as explainability, that cross application boundaries and deserve a different level of attention. Research libraries can be influential in these debates given their expertise in knowledge management and research support, and their concern for the public good.

An interesting example arises in the area of copyright as a result of discussion about the ownership of materials created by an AI. This has led some to argue for the creation of "AI sunshine laws," which would mirror the idea of the public domain in copyright or patent law. The code and logic of the AI system would, at some point, become public, transparent, and open to scrutiny and reuse. This requirement would position AI within more traditional IP legislation and would extend the notion of public domain into new and likely highly contentious areas.

Algorithmic Literacy

Research libraries, like all libraries, have been active proponents of enhancing literacy, be it traditional reading and writing or more recently digital literacy in all its various forms. While algorithmic literacy can be seen as a subset of digital literacy or computational thinking, it has unique characteristics and applications that deserve specific attention. Just as information literacy provides users with skills and perspectives to assess resources, algorithmic literacy is an explainability strategy allowing users to navigate and utilize algorithmic tools and services.

Calling "algorithmic awareness" a "new competency," the objective of the 2017 Institute of Museum and Library Services (IMLS) grant proposal from Jason Clark and colleagues at Montana State University is to "find transparency for the invisible logic embedded in our software interactions. Success in this setting would be our community finding new teaching methods and confidence to make this logic visible for our patrons and ourselves."[27] By linking algorithmic awareness to

> Just as information literacy provides users with skills and perspectives to assess resources, algorithmic literacy is an explainability strategy allowing users to navigate and utilize algorithmic tools and services.

information and digital literacy, Clark identifies a gap in the Association of College & Research Libraries information literacy framework revealing "a lack of an understanding around the rules that govern our software and shape our digital experiences."[28] The anticipated "Algo Report" from Project Information Literacy will present findings from a national study of college students in the US and address "how algorithms affect the information that streams at them constantly throughout the day in order to be truly information literate in the 21st century."[29]

AI-Authorship: An Explainability Sandbox

An interesting and instructive example of the role of XAI in research libraries arose earlier this year when Springer Nature published an open access book written by AI: *Lithium-Ion Batteries: A Machine-Generated Summary of Current Research*.[30] The author, identified as "Beta Writer," algorithmically categorized and summarized more than 150 key research publications selected from over 1,000 published from 2016 to 2018, thereby synthesizing a large and complex corpus of the current research literature. The algorithmic processes that created this book, using a combination of various "off the shelf" natural language processing (NLP) tools, included preprocessing the documents to address various linguistic and semantic normalizations; clustering documents by content similarity (that is, the content in the chapters and sections of the book); generating abstracts, summaries, introductions, and conclusions; and finally outputting XML as a completed manuscript.

The details are outlined in a human-written preface and provide an interesting comparison to current cataloging and metadata processes and to accepted scholarly communication practices.[31] Henning

Schoenenberger, director of product data and metadata management at Springer Nature, is clear that the intent of the project is "to initiate a public debate on the opportunities, implications and potential risks of machine-generated content in scholarly publishing."[32] Springer has gone to great lengths to document their process, discuss alternative strategies, identify weaknesses and outright failures, and to encourage critical commentary. In many ways they have provided an "explainability sandbox" for scholarly publishing. Determining the value of this and similar books will be achieved in part by interrogating the methods and processes by which they are constructed. In other words, the emerging AI books will need the capacity to explain themselves.

Conclusion

In his article about stewardship in the "age of algorithms," Clifford Lynch argues that algorithmic accountability is "the domain of the regulator or social justice advocate, not the archivist."[33] However, he also notes that "this new world is strange and inhospitable to most traditional archival practice" and that "our thinking about a good deal of the digital world must shift from artifacts requiring mediation and curation, to experiences."[34] These observations suggest that the role of the archivist (and of research libraries more generally) should indeed include algorithmic accountability because of its centrality to emerging practices.

> "Research libraries have a unique and important opportunity to shape the development, deployment, and use of intelligent systems in a manner consistent with the values of scholarship and librarianship."

The complexity and opacity of algorithmic decision-making, replete with limitations, outright failures, and dramatic advances, is challenging and changing our notions of information systems and their use. The field of explainable AI has emerged as a set of strategies, techniques, and processes used in a variety of contexts to facilitate trust and accountability. As key stakeholders in the scholarly communications ecosystem being significantly disrupted by artificial intelligence, research libraries have a unique and important opportunity to shape the development, deployment, and use of intelligent systems in a manner consistent with the values of scholarship and librarianship. The area of explainable artificial intelligence is only one component of this, but in many ways, it may be the most important.

Endnotes

1. Safiya Umoja Noble, *Algorithms of Oppression: How Search Engines Reinforce Racism* (New York: NYU Press, 2018), 1.

2. Frank Pasquale, *The Black Box Society: The Secret Algorithms That Control Money and Information* (Cambridge, MA: Harvard University Press, 2015), 8.

3. Jos de Mul and Bibi van den Berg, "Remote Control: Human Autonomy in the Age of Computer-Mediated Agency," in *Law, Human Agency and Autonomic Computing*, ed. Mireille Hildebrandt and Antoinette Rouvroy (Abingdon, UK: Routledge, 2011).

4. Pasquale, *Black Box Society*, 18.

5. Mariarosaria Taddeo, "Trusting Digital Technologies Correctly," *Minds and Machines*, 27 no. 4 (2017): 565–568, https://doi.org/10.1007/s11023-017-9450-5.

6. *Explainable Artificial Intelligence (XAI)*, Broad Agency Announcement DARPA-BAA-16-53 (Arlington, VA: DARPA, August 10, 2016), 5, http://www.darpa.mil/attachments/DARPA-BAA-16-53.pdf.

7. Matt Turek, "Explainable Artificial Intelligence (XAI)," DARPA, accessed August 30, 2019, https://www.darpa.mil/program/explainable-artificial-intelligence.

8. Siddhartha Mukherjee, "The Algorithm Will See You Now," *New Yorker*, April 3, 2017, 46–53, https://www.newyorker.com/magazine/2017/04/03/ai-versus-md.

9. Tania Lombrozo, "The Structure and Function of Explanations," *Trends in Cognitive Sciences* 10, no. 10 (2006): 464–470, https://doi.org/10.1016/j.tics.2006.08.004.

10. Emilee Rader, Kelley Cotter, and Janghee Cho, "Explanations as Mechanisms for Supporting Algorithmic Transparency," *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems* (New York: ACM, 2018), https://doi.org/10.1145/3173574.3173677.

11. Dino Pedreschi et al., "Open the Black Box: Data-Driven Explanation of Black Box Decision Systems," preprint, submitted June 26, 2018, http://arxiv.org/abs/1806.09936.

12. Taina Bucher, *If...Then: Algorithmic Power and Politics* (New York: Oxford University Press, 2018).

13. Tim Miller, "Explanation in Artificial Intelligence: Insights from the Social Sciences," *Artificial Intelligence*, 267 (2019): 1–38, https://doi.org/10.1016/j.artint.2018.07.007.

14. Cliff Kuang, "Can A.I. be Taught to Explain Itself?," *New York Times Magazine*, November 21, 2017, https://nyti.ms/2hR1S15.

15. Alan F. T. Winfield and Marina Jirotka, "The Case for an Ethical Black Box," in *Towards Autonomous Robotic Systems*, ed. Yang Gao, Saber Fallah, Yaochu Jin, and Constantina Lekakou, Lecture Notes in Computer Science 10454 (Cham, Switzerland: Springer Nature, 2017), 262, https://doi.org/10.1007/978-3-319-64107-2_21.

16. Winfield and Jirotka, 269.

17. Amitai Etzioni and Oren Etzioni, "Keeping AI Legal," *Vanderbilt Journal of Entertainment & Technology Law* 19, no. 1 (2016): 136, https://doi.org/10.2139/ssrn.2726612.

18. Joelle Pineau, "ICLR Reproducibility Challenge: Second Edition, 2019," Pineau's website, accessed August 30, 2019, https://www.cs.mcgill.ca/~jpineau/ICLR2019-ReproducibilityChallenge.html.

19. Lee Rainie and Janna Anderson, *Code-Dependent: Pros and Cons of the Algorithm Age* (Washington, DC: Pew Research Center, February 2017), http://www.pewinternet.org/wp-content/uploads/sites/9/2017/02/PI_2017.02.08_Algorithms_FINAL.pdf.

20. R. David Lankes, "Decoding AI and Libraries," Lankes's website, July 3, 2019, https://davidlankes.org/decoding-ai-and-libraries/.

21. Catherine Nicole Coleman, "Artificial Intelligence and the Library of the Future, Revisited," *Digital Library Blog*, Stanford Libraries, November 3, 2017, http://library.stanford.edu/blogs/digital-library-blog/2017/11/artificial-intelligence-and-library-future-revisited.

22. "Artificial Intelligence and Machine Learning in Libraries," ed. Jason Griffey, *Library Technology Reports* 55, no. 1 (2019), https://doi.org/10.5860/ltr.55n1.

23. Chris Bourg, "What Happens to Libraries and Librarians When Machines Can Read All the Books?," *Feral Librarian*, March 16, 2017, https://chrisbourg.wordpress.com/2017/03/16/what-happens-to-libraries-and-librarians-when-machines-can-read-all-the-books/.

24. Vahe Tshitoyan et al., "Unsupervised Word Embeddings Capture Latent Knowledge from Materials Science Literature," *Nature* 571, no. 7763 (2019): 95–98, https://doi.org/10.1038/s41586-019-1335-8.

25. Don R. Swanson, "Medical Literature as a Potential Source of New Knowledge," *Bulletin of the Medical Library Association* 78, no. 1 (January 1990): 29–37, https://www.ncbi.nlm.nih.gov/pmc/articles/PMC225324/.

26. Jacobo Elosua, Anita Schjøll Brede, Maria Ritola, and Victor Botev, "Iris.ai's Project Aiur: An Open, Community-Governed AI Engine for Knowledge Validation," Iris.ai, May 2018, https://iris.ai/wp-content/uploads/2018/05/ProjectAiur_whitepaper.pdf.

27. Jason Clark, Lisa Janicke Hinchliffe, and Scott Young, "'RE:Search'—Unpacking the Algorithms That Shape Our UX," Institute of Museum and Library Services, 2017, https://www.imls.gov/sites/default/files/grants/re-72-17-0103-17/proposals/re-72-17-0103-17-full-proposal-documents.pdf.

28. Jason Clark, "Building Competencies around Algorithmic Awareness" (presentation at Code4Lib, Washington, DC, 2018), Clark's website, https://www.lib.montana.edu/~jason/talks/algorithmic-awareness-talk-code4lib2018.pdf.

29. Alison J. Head and Barbara Fister, "The Algo Report 2019," Project Information Literacy, March 28, 2019, https://www.projectinfolit.org/uploads/2/7/5/4/27541717/about-_pils_algo_report-2019.pdf.

30. Beta Writer, *Lithium-Ion Batteries: A Machine-Generated Summary of Current Research* (Heidelberg: Springer, 2019), https://link.springer.com/book/10.1007/978-3-030-16800-1.

31. Henning Schoenenberger, Christian Chiarcos, and Niko Schenk, "Preface," in Beta Writer, *Lithium-Ion Batteries: A Machine-Generated Summary of Current Research* (Heidelberg: Springer, 2019) https://link.springer.com/content/pdf/10.1007%2F978-3-030-16800-1.pdf.

32. Lettie Y. Conrad, "The Robots Are Writing: Will Machine-Generated Books Accelerate Our Consumption of Scholarly Literature?," *Scholarly Kitchen*, June 25, 2019, https://scholarlykitchen.sspnet.org/2019/06/25/the-robots-are-writing-will-machine-generated-books-accelerate-our-consumption-of-scholarly-literature/.

33. Clifford Lynch, "Stewardship in the 'Age of Algorithms'," *First Monday* 22, no. 12 (December 2017), http://firstmonday.org/ojs/index.php/fm/article/view/8097.

34. Lynch.

To cite this article: Michael Ridley. "Explainable Artificial Intelligence." *Research Library Issues*, no. 299 (2019): 28–46. https://doi.org/10.29242/rli.299.3.